BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION
Statistical Research Report Series
No. RR2001/02

Convergence of a Robbins-Monro Algorithm for Recursive Parameter Estimation with Non-Monotone Weights and Multiple Zeros

David F. Findley
Statistical Research Division
Methodology and Standards Directorate
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: June 26, 2001

Convergence of a Robbins-Monro Algorithm for Recursive Parameter Estimation with Non-Monotone Weights and Multiple Zeroes

David F. Findley
U.S. Census Bureau

June 7, 2001

Abstract

The convergence of the output of a deterministic recursive algorithm with transient errors is proved. This result is mostly implicit in the proof of a stochastic approximation result by Fradkov that has never been translated into English. A gap in Fradkov's proof is fixed but only for the scalar case. Fradkov's restriction to monotonically decreasing weight is avoided.

In this note, we present a result on the convergence of a deterministic algorithm of Robbins-Monro form ((4) below) with transient error ((3) below). In its essentials, this result has been extracted from the proof of Theorem 3.17 of the monograph by Derevitzkiĭ and Fradkov (1981) (hereafter DF) concerning the almost sure convergence of a stochastic approximation scheme. In DF, this theorem and the sequence of Lemmas P.12–P.16 that constitute its proof are credited to Fradkov. Although formulated differently, Lemma P.12 can be interpreted as reducing the proof the convergence of the stochastic approximation method considered in Theorem 3.17 to the proof of a result like the Proposition below.

Our Proposition avoids the hypothesis of Fradkov's theorem that the weighting sequence δ_t decreases monotonically. DF explicitly refer to this hypothesis only in the proof of Lemma P.12 and there is only one place in the proof of the subsequent lemmas where additional discussion is needed if monotonicity is not assumed. However, there is a gap in the proof of the main supporting result, Lemma P. 16, that cannot be bridged without additional restrictions except in the case of sequences θ_t of dimension $d = 1$. For this reason, the Proposition and the final lemma of its proof, Lemma 5, are restricted to scalar case. The gap is described after the proof of Lemma 5. Although the sequence of lemmas and proofs given below follows closely the sequence of lemmas and proofs used by DF, our presentation provides greater precision in both statements and proofs of the lemmas. Because an English translation of DF is not available, and because our assumptions (2) are the least restrictive assumptions possible for δ_t, which significantly increases the applicability of the Proposition, it seemed worthwhile to give complete details as we do in this note. An attractive feature of the Proposition is that it can be applied to the analysis of recursively estimated time series model parameters in the situation of a misspecified model. In this situation parameter estimates need not converge to unique limits, see Section 4.3 of Findley, Pötscher, and Wei (2001) for a survey of the relevant literature. Some rigorous convergence results for time series model parameter recursions whose proofs utilize the Proposition can be found in Cantor (2001).

Because only the final lemma used prove the Proposition requires the recursively defined sequence θ_t to be scalar, we shall use a general notation to obtain full generality in the statements and proofs of the preceding lemmas. For a d-dimensional column vector $x = (x_1, \ldots, x_d)^T$, we define $\|x\| = \left(\sum_{i=1}^d x_i^2\right)^{1/2}$. For a $d \times d$ matrix M, we define $\|M\| = \lambda_{\max}^{1/2}\left(M^T M\right)$, where $\lambda_{\max}$ denotes the largest eigenvalue.

PROPOSITION. *Let Θ denote an open subset of the space R^d of d-dimensional real column vectors. Let $f\colon \Theta \longmapsto R^d$ be a continuously differentiable function on Θ with the following properties:*

(a) The set $\Theta_0 = \{\theta \in \Theta : f(\theta) = 0\}$ is nonempty and compact.

(b) There is a bounded open subset Θ_V of Θ containing Θ_0 on which there is a twice continuously differential function $V\colon \Theta_V \longmapsto R$ such that, for all $\theta \notin \Theta_0$ in Θ_V, the derivative $\nabla V = (\partial V/\partial\theta_1, \ldots, \partial V/\partial\theta_d)^T$ has the property

$$\nabla V(\theta)^T f(\theta) > 0. \tag{1}$$

Suppose that for some sequence of real numbers $\delta_t, t \geq 1$ satisfying

$$\delta_t \geq 0, \lim_{t\to\infty} \delta_t = 0, \sum_{t=0}^{\infty} \delta_t = \infty, \tag{2}$$

and for some R^d-valued sequence w_t satisfying

$$\lim_{t\to\infty} w_t = 0, \tag{3}$$

the sequence θ_t, $t \geq 1$ in Θ satisfies

$$\theta_t = \theta_{t-1} - \delta_t f\left(\theta_{t-1}\right) + \delta_t w_t \tag{4}$$

for $t \geq 2$ and enters Θ_V infinitely often but has no cluster point on the boundary $\partial\Theta_V$ of Θ_V.

Under these conditions, when $d = 1$, *every cluster point of θ_t belongs to $\Theta_0 = \{\theta \in \Theta : f(\theta) = 0\}$. That is,* as $t \to \infty$,

$$\theta_t \to \Theta_0. \tag{5}$$

In particular, if $\Theta_0 = \{\theta_\infty\}$, then $\lim_{t\to\infty} \theta_t = \theta_\infty$.

The proof is obtained via a sequence of lemmas and the following observations. First, the assumptions of the Proposition yield that:

(i) There are only finitely many t' such that $\theta_{t'} \in \Theta_V$ but $\theta_{t'+1} \notin \Theta_V$. Otherwise, since Θ_V is bounded, a subsequence $\theta_{t''}$ of $\theta_{t'}$ would have a limit $\overline{\theta}$, with $\overline{\theta} \notin \partial\Theta_V$, by assumption, so $\overline{\theta} \in \Theta_V$. By continuity, $\lim_{t''\to\infty} f\left(\theta_{t''}\right) = f\left(\overline{\theta}\right)$, and (4),(2), and (3) would yield $\lim_{t''\to\infty} \{\theta_{t''+1} - \theta_{t''}\} = 0$, and therefore $\lim_{t''\to\infty} \theta_{t''+1} = \overline{\theta}$, and hence that $\theta_{t''+1} \in \Theta_V$ for t'' sufficiently large (because Θ_V is open), which contradicts the definition of the $\theta_{t'}$ sequence. Consequently, there is a time t_V such that

$$\theta_t \in \Theta_V,\ t \geq t_V. \tag{6}$$

For the subsequent discussion, we shall always assume that $t \geq t_V$.

(ii) It follows from (6) that the sequence θ_t is bounded. Hence its set of cluster points,

$$K = \left\{\overline{\theta} : \overline{\theta} = \lim_{t'\to\infty} \theta_{t'} \text{ for some subsequence } \theta_{t'} \text{ of } \theta_t\right\},$$

is nonempty and compact. Since $K \cap \partial\Theta_V$ is empty, $K \subset \Theta_V$. Consequently, there exist $\rho, \zeta > 0$ depending only on K such that for each $\overline{\theta} \in K$, the closed ball $B(\overline{\theta}, 2\rho) = \{\theta \in \Theta_V : \|\theta - \theta\| \leq 2\rho\}$ has the properties that

$$B(\overline{\theta}, 2\rho) \subseteq \Theta_V, \tag{7}$$

holds, and, for any point θ^* on the boundary $\partial\Theta_V$ of Θ_V, also

$$\min_{\theta \in B(\overline{\theta}, 2\rho)} \|\theta - \theta^*\| \geq \zeta. \tag{8}$$

It follows from (8) that the closure $\tilde{K} = \tilde{K}(\rho)$ of $\cup_{\overline{\theta} \in K} B(\overline{\theta}, 2\rho)$ is a compact subset of Θ_V containing K in its interior, $K \subset Int\tilde{K}$. By the continuity of $\nabla f(\theta)$, $L = \max_{\theta \in \tilde{K}} \|\nabla f(\theta)\|$ is finite. Hence, from the Mean Value Theorem and the convexity of $B(\overline{\theta}, 2\rho)$, and we have

$$\|f(\theta) - f(\theta')\| \leq L \|\theta - \theta'\| \tag{9}$$

for every $\theta, \theta' \in B(\overline{\theta}, 2\rho)$ when $\overline{\theta} \in K$.

For any sequence δ_t satisfying (2), we define, for every $\Delta > 0$,

$$t_\Delta = \min \{t_0 : \delta_t + \delta_{t+1} \leq \Delta \text{ for all } t \geq t_0\} \tag{10}$$

and, for every $t \geq t_\Delta$

$$u_\Delta(t) = \max \{u \geq t : \delta_t + \cdots + \delta_u \leq \Delta\}. \tag{11}$$

Because $\sum_{t=0}^{\infty} \delta_t = \infty$, we always have $u_\Delta(t) < \infty$. Also, $u_\Delta(t) \geq t + 1$.

Lemma 1 . *For any sequence δ_t satisfying (2) and any $\Delta > 0$, we have*

$$\lim_{t \to \infty} \left\{ \Delta - \sum_{u=t}^{u_\Delta(t)} \delta_u \right\} = 0. \tag{12}$$

Proof. If not, there would exist $\Delta, \varepsilon > 0$ and a subsequence $\delta_{t'}$ such that $\Delta - \sum_{u=t'}^{u_\Delta(t')} \delta_u \geq 2\varepsilon$ held for all t'. But if we choose t_ε so that $\delta_t \leq \varepsilon$ whenever $t \geq t_\varepsilon$, then for $t' \geq \max\{t_\varepsilon, t_\Delta\}$, we would have $\Delta - \sum_{u=t'}^{u_\Delta(t')+1} \delta_u \geq \varepsilon > 0$, contradicting the definition of $u_\Delta(t')$.

Lemma 2 . *(cf. Lemma P.13 of DF). For each $\rho > 0$ such that (7) and (8) hold, there exists a $\Delta_0 = \Delta_0(\rho)$ such that for every $0 < \Delta \leq \Delta_0$ and any $\overline{\theta} \in K$, if*

$$\left\|\theta_t - \overline{\theta}\right\| \leq \rho, \tag{13}$$

holds, then so does

$$\sup_{t \leq u \leq u_\Delta(t)} \left\|\theta_u - \overline{\theta}\right\| \leq 2\rho \tag{14}$$

for every $t \geq t_\Delta$, for t_Δ and $u_\Delta(t)$ defined in (10) and (11).

Proof. We use induction to establish (14). Let $t \leq v < u_\Delta(t)$ be such that

$$\sup_{t \leq u \leq v} \left\| \theta_u - \overline{\theta} \right\| \leq 2\rho. \tag{15}$$

(This holds for $v = t$ from (13).) From (4),

$$\begin{aligned} \theta_{v+1} &= \theta_t - \sum_{u=t}^{v} \delta_{u+1} f(\theta_u) + \sum_{u=t}^{v} \delta_{u+1} w_{u+1} \\ &= \theta_t - \sum_{u=t}^{v} \delta_{u+1} f\left(\overline{\theta}\right) + \sum_{u=t}^{v} \delta_{u+1} \left\{ f\left(\overline{\theta}\right) - f(\theta_u) \right\} + \sum_{u=t}^{v} \delta_{u+1} w_{u+1}. \end{aligned} \tag{16}$$

Note that with $r_\Delta(t) = \Delta \max_{u \geq t} |w_u|$, we have

$$\max_{t \leq v \leq u_\Delta(t) - 1} \left\| \sum_{u=t}^{v} \delta_{u+1} w_{u+1} \right\| \leq r_\Delta(t). \tag{17}$$

It follows from (16) and (9) that

$$\left\| \theta_{v+1} - \overline{\theta} \right\| \leq \left\| \theta_t - \overline{\theta} \right\| + \Delta \left\| f\left(\overline{\theta}\right) \right\| + r_\Delta(t) + L \sum_{u=t}^{v} \delta_{u+1} \left\| \theta_u - \overline{\theta} \right\|.$$

Therefore, from an induction argument or from the Discrete Bellman-Gronwall Lemma (Solo and Kong, 1995 p. 315) and the fact that $e^x \geq 1 + x$ for any $x \geq 0$, we have

$$\begin{aligned} \left\| \theta_{v+1} - \overline{\theta} \right\| &\leq \left\{ \left\| \theta_t - \overline{\theta} \right\| + \Delta \left\| f\left(\overline{\theta}\right) \right\| + r_\Delta(t) \right\} \prod_{u=t}^{v} (1 + L\delta_{u+1}) \\ &\leq \left\{ \left\| \theta_t - \overline{\theta} \right\| + \Delta \left\| f\left(\overline{\theta}\right) \right\| + r_\Delta(t) \right\} e^{L \sum_{u=t}^{v} \delta_{u+1}} \\ &\leq \left\{ \left\| \theta_t - \overline{\theta} \right\| + \Delta \left(\left\| f\left(\overline{\theta}\right) \right\| + \max_{u \geq t} |w_u| \right) \right\} e^{L\Delta}. \end{aligned} \tag{18}$$

Define $\Delta_0(\rho)$ to be the largest Δ_0 for which

$$L\Delta_0 \leq \log \frac{4}{3},\ \Delta_0 \left(\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| + \max_{u \geq 1} |w_u| \right) \leq \rho/2 \tag{19}$$

holds. Then for each $\Delta \leq \Delta_0(\rho)$ and $t \geq t_\Delta$ for which (13) and (15) hold, it follows from (18) and (13) that $\left\| \theta_{v+1} - \overline{\theta} \right\| \leq 2\rho$. Thus, by induction (14) holds for all $t \geq t_\Delta, \Delta \leq \Delta_0(\rho)$ for which (13) holds, as asserted.

Lemma 3 *(cf. Lemma P.14 of DF). For $\overline{\theta} \in K$, ρ as in (7) and Δ_0 as in (19), and any $0 < \Delta \leq \Delta_0$ and $t \geq t_\Delta$ such that (13) holds, we have*

$$\theta_{u_\Delta(t)} = \theta_t - f\left(\overline{\theta}\right)\Delta + q_1(t,\Delta) + q_2(t,\Delta), \tag{20}$$

where $q_1(t,\Delta)$ has the property that there exist constants C_1, C_2 such that

$$\|q_1(t,\Delta)\| \leq C_1\Delta\left\|\theta_t - \overline{\theta}\right\| + C_2\Delta^2, \tag{21}$$

and $q_2(t,\Delta)$ satisfies

$$\lim_{t\to\infty} q_2(t,\Delta) = 0. \tag{22}$$

Proof. From (4), we obtain

$$\begin{aligned}\theta_{u_\Delta(t)} &= \theta_t - f\left(\overline{\theta}\right)\Delta + \left(\Delta - \sum_{u=t}^{u_\Delta(t)-1}\delta_{u+1}\right) f\left(\overline{\theta}\right) \\ &\quad + \sum_{u=t}^{u_\Delta(t)-1}\delta_{u+1}\left\{f\left(\overline{\theta}\right) - f(\theta_u)\right\} + \sum_{u=t}^{u_\Delta(t)-1}\delta_{u+1}w_{u+1}.\end{aligned}$$

Set

$$q_1(t,\Delta) = \sum_{u=t}^{u_\Delta(t)-1}\delta_{u+1}\left\{f\left(\overline{\theta}\right) - f(\theta_u)\right\},$$

and

$$q_2(t,\Delta) = \left(\Delta - \sum_{u=t}^{u_\Delta(t)-1}\delta_{u+1}\right) f\left(\overline{\theta}\right) + \sum_{u=t}^{u_\Delta(t)-1}\delta_{u+1}w_{u+1}. \tag{23}$$

From (9) we have

$$\|q_1(t,\Delta)\| \leq L\sum_{u=t}^{u_\Delta(t)-1}\delta_{u+1}\left\|\theta_u - \overline{\theta}\right\|, \tag{24}$$

and from (18) and (19),

$$\sum_{u=t}^{u_\Delta(t)-1}\delta_{u+1}\left\|\theta_u - \overline{\theta}\right\| = \delta_{t+1}\left\|\theta_t - \overline{\theta}\right\| + \sum_{u=t+1}^{u_\Delta(t)-1}\delta_{u+1}\left\|\theta_u - \overline{\theta}\right\|$$

$$\leq \delta_{t+1} \left\| \theta_t - \overline{\theta} \right\|$$
$$+ \left\{ \sum_{u=t+1}^{u_\Delta(t)-1} \delta_{u+1} \left\| \theta_t - \overline{\theta} \right\| + \left(\sum_{u=t+1}^{u_\Delta(t)-1} \delta_{u+1} \right) \Delta \left(\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| + \max_{u \geq t} |w_{u+1}| \right) \right\} e^{L\Delta}$$
$$\leq \frac{4}{3} \Delta \left\| \theta_t - \overline{\theta} \right\| + \frac{4}{3} \Delta^2 \left(\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| + \max_{u \geq 1} |w_u| \right).$$

This yields (21) with

$$C_1 = \frac{4}{3} L, \quad C_2 = \frac{4}{3} L \left(\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| + \max_{u \geq 1} |w_u| \right). \tag{25}$$

The assertion (22) concerning $q_2(t, \Delta)$ follows from Lemma 1, $\delta_t \to 0$, $\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| < \infty$, and

$$\lim_{t \to \infty} \left\| \sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1} w_{u+1} \right\| \leq \Delta \lim_{t \to \infty} \max_{u \geq t+1} \|w_u\| = 0,$$

by virtue of (17) and (3).

The final two lemmas use the properties of $V(\theta)$.

Lemma 4 *(cf. Lemma P.15 of DF). Suppose $\overline{\theta} \in K$ is such that $f\left(\overline{\theta}\right) \neq 0$. Then for each subsequence $\theta_{t'}$ converging to $\overline{\theta}$, there exist $\Delta_0 > 0$ and $\eta > 0$ with the following property: for each $0 < \Delta \leq \Delta_0$ there is a $t'(\Delta)$ such that for all $t' \geq t'(\Delta)$, the inequality*

$$V\left(\theta_{u_\Delta(t')}\right) < V(\theta_{t'}) - \eta\Delta \tag{26}$$

holds.

Proof. Let $\rho > 0$ be such that (7) and (8) hold. For each $0 < \varepsilon < \rho$, let $\Delta_0(\varepsilon)$ be the largest Δ_0 for which (19) holds when ρ is replaced by ε. For $\Delta \leq \Delta_0(\varepsilon)$ consider a $t' \geq t_\Delta$ such that $\left\| \theta_{t'} - \overline{\theta} \right\| \leq \varepsilon$ holds, and therefore $\left\| \theta_{u_\Delta(t')} - \overline{\theta} \right\| \leq 2\varepsilon$ by Lemma 2. To simplify notation, set $\theta' = \theta_{t'}$ and $\theta'' = \theta_{u_\Delta(t')}$. By taking Taylor expansions of V and ∇V, we obtain

$$V(\theta'') - V(\theta') = \nabla V(\zeta)^T (\theta'' - \theta')$$

$$= \nabla V\left(\overline{\theta}\right)^T\left(\theta''-\theta'\right)+\left[\nabla V\left(\zeta\right)-\nabla V\left(\overline{\theta}\right)\right]^T\left(\theta''-\theta'\right)$$

$$= \nabla V\left(\overline{\theta}\right)^T\left(\theta''-\theta'\right)+\left(\zeta-\overline{\theta}\right)^T\nabla^2 V\left(\zeta'\right)\left(\theta''-\theta'\right) \tag{27}$$

with $\zeta \in [\theta', \theta'']$ and $\zeta' \in \left[\zeta, \overline{\theta}\right]$. ($[\theta', \theta''] = \{\alpha\theta' + (1-\alpha)\theta'' : 0 \le \alpha \le 1\}$, etc.) Since $B\left(\overline{\theta}, 2\varepsilon\right)$ is convex, $\zeta, \zeta' \in B\left(\overline{\theta}, 2\varepsilon\right)$. From Lemma 3,

$$\theta'' - \theta' = -f\left(\overline{\theta}\right)\Delta + q_1\left(t', \Delta\right) + q_2\left(t', \Delta\right), \tag{28}$$

where

$$\left\|q_1\left(t, \Delta\right)\right\| \le C_1\Delta\varepsilon + C_2\Delta^2$$

with C_1, C_2 given by (25), and where $\lim_{t'\to\infty} q_2\left(t', \Delta\right) = 0$. Since $\overline{\theta} \notin \Theta_0$, it follows from (1) that $\nabla V\left(\overline{\theta}\right)^T f\left(\overline{\theta}\right) = \eta_1 > 0$. Let η satisfy $0 < \eta < \eta_1$ and set $\tilde{\eta} = \eta_1 - \eta$. Substituting (28) into (27), we obtain

$$\begin{aligned} V\left(\theta''\right) - V\left(\theta'\right) &= -\eta_1\Delta \\ &- \left(\zeta-\overline{\theta}\right)^T\nabla^2 V\left(\zeta'\right) f\left(\overline{\theta}\right)\Delta \\ &+ \left[\nabla V\left(\overline{\theta}\right)^T + \left(\zeta-\overline{\theta}\right)^T\nabla^2 V\left(\zeta'\right)\right] q_1\left(t', \Delta\right) \\ &+ \left[\nabla V\left(\overline{\theta}\right)^T + \left(\zeta-\overline{\theta}\right)^T\nabla^2 V\left(\zeta'\right)\right] q_2\left(t', \Delta\right). \end{aligned} \tag{29}$$

Set

$$L_1 = \max_{\theta\in\tilde{K}}\left\|\nabla V\left(\theta\right)\right\|, \quad L_2 = \max_{\theta\in\tilde{K}}\left\|\nabla^2 V\left(\theta\right)\right\| \tag{30}$$

and $C_3 = L_1 + 2\rho L_2$. Now choose ε small enough that

$$2\varepsilon L_2 \max_{\theta\in\tilde{K}}\left|f\left(\theta\right)\right| < \frac{\tilde{\eta}}{3}, \;\; C_1C_3\varepsilon < \frac{\tilde{\eta}}{6}$$

and also so that

$$C_2C_3\Delta_0\left(\varepsilon\right) < \frac{\tilde{\eta}}{6}.$$

Then, for any $0 < \Delta \le \Delta_0\left(\varepsilon\right)$, if we choose $t'\left(\Delta\right) \ge t_\Delta$ so that $t \ge t'\left(\Delta\right)$ implies

$$\left\|\theta_{t'} - \overline{\theta}\right\| \le \varepsilon$$

and

$$C_3 q_2 (t, \Delta) < \frac{\tilde{\eta}}{3} \Delta,$$

it follows from (29) that

$$V (\theta'') < V (\theta') - \eta_1 \Delta + \tilde{\eta} \Delta = V (\theta') - \eta \Delta,$$

holds when $t' \geq t' (\Delta)$, as asserted.

The final lemma is the first to require $d = 1$. Under this condition, for every $\overline{\theta} \in \Theta_V$ for which $f \left(\overline{\theta} \right) \neq 0$, it follows from (1), the Mean Value Theorem, and the continuity of $\nabla V (\theta)$ that there exist $m > 0, \rho > 0$ such that

$$\left| V (\theta) - V \left(\overline{\theta} \right) \right| \geq m \left\| \theta - \overline{\theta} \right\| \tag{31}$$

holds for all $\theta \in B \left(\overline{\theta}, 2\rho \right)$.

Lemma 5 *(cf. Lemma P.16 of DF). Under the assumptions of the Proposition, no point $\overline{\theta} \in \Theta_V$ for which $f \left(\overline{\theta} \right) \neq 0$ can be a limit point of θ_t. Therefore $K \subseteq \Theta_0$.*

Proof. Suppose, to the contrary, that there is a $\overline{\theta} \in \Theta_V$ with $f \left(\overline{\theta} \right) \neq 0$ (and therefore with $\eta_1 = \nabla V \left(\overline{\theta} \right)^T f \left(\overline{\theta} \right) > 0$) which is a limit point of θ_t. Choose ρ so that (7) and (8) are satisfied, and also so that

$$f (\theta) \neq 0 \tag{32}$$

and (31) hold for all $\theta \in B \left(\overline{\theta}, 2\rho \right)$. For Δ_0 as in the proof of Lemma 4, choose $\Delta, \eta_0 > 0$ so that

$$\Delta \leq \min \{ \Delta_0, \rho m \}, \ \ \eta_0 < \min \{ \eta_1, 1 \} . \tag{33}$$

Then Lemma 4's proof shows that, for any subsequence $\theta_{t'}$ that converges to $\overline{\theta}$, the inequality

$$V \left(\theta_{u_\Delta (t')} \right) < V (\theta_{t'}) - \eta_0 \Delta \tag{34}$$

holds for all t' large enough. The sequence of values $\theta_{u_\Delta (t')}$ appearing on the l.h.s. of (34) does not necessarily change with t'. Since δ_t need not be monotonically decreasing, all that can be asserted is that for $t'' > t' \geq t_\Delta$, one has $u_\Delta(t'') \geq u_\Delta(t')$, with $u_\Delta(t'') > u_\Delta(t')$ holding for $t'' \geq u_\Delta(t')$. The latter inequality guarantees that $\theta_{u_\Delta(t')}$ takes on infinitely many values of θ_t. From this fact and (34), and from $V(\theta_{t'}) \to V(\overline{\theta})$, we can conclude that, for a given $0 < \eta < \eta_0$, the sequence θ_t enters each of the disjoint sets

$$R_{\eta\Delta} = \left\{\theta \in B\left(\overline{\theta}, 2\rho\right) : V(\theta) \leq V\left(\overline{\theta}\right) - \eta\Delta\right\}$$

and

$$S_{\frac{1}{2}\eta\Delta} = \left\{\theta \in B\left(\overline{\theta}, 2\rho\right) : V(\theta) > V\left(\overline{\theta}\right) - \frac{\eta}{2}\Delta\right\}$$

infinitely often. Let $\theta_{\tau'}$ denote the subsequence of last values of θ_t in $R_{\eta\Delta}$ before a next entry $\theta_{\tau'+n'}$ in $S_{\frac{1}{2}\eta\Delta}$. Some subsequence $\theta_{\tau''}$ of $\theta_{\tau'}$ must have a limit $\tilde{\theta}$. Since $V(\theta_{\tau''+1}) > V\left(\overline{\theta}\right) - \eta\Delta$ and $V(\theta_{\tau''+1}) - V(\theta_{\tau''}) \to 0$, we must have $V\left(\tilde{\theta}\right) = V\left(\overline{\theta}\right) - \eta\Delta$. Therefore, from (31) and (33),

$$\left\|\tilde{\theta} - \overline{\theta}\right\| < \rho. \tag{35}$$

Thus, $f\left(\tilde{\theta}\right) \neq 0$ by (32). With $\Delta_0(\rho)$ as in Lemma 1 and L_1 as in (30), we can conclude from Lemma 4 that there exist $\widetilde{\Delta} > 0$ satisfying

$$\widetilde{\Delta} < \min\left\{\frac{\rho}{2}, \frac{\eta}{6L_1}\Delta, \Delta_0(\rho)\right\}, \tag{36}$$

and $0 < \tilde{\eta} < 1$ and $\tau''\left(\widetilde{\Delta}\right) \geq t_{\widetilde{\Delta}}$ such that $V\left(\theta_{u_{\widetilde{\Delta}}(\tau'')}\right) < V(\theta_{\tau''}) - \tilde{\eta}\widetilde{\Delta}$ holds for all $\tau'' \geq \tau''\left(\widetilde{\Delta}\right)$. Hence

$$V\left(\theta_{u_{\widetilde{\Delta}}(\tau'')}\right) < V\left(\overline{\theta}\right) - \eta\Delta - \tilde{\eta}\widetilde{\Delta}. \tag{37}$$

Because $\theta_{\tau''} \to \tilde{\theta}$, we can, by taking a larger $\tau''\left(\widetilde{\Delta}\right)$ if necessary, further obtain

$$\left\|\theta_{\tau''} - \tilde{\theta}\right\| \leq \widetilde{\Delta}, \tag{38}$$

for all $\tau'' \geq \tau''\left(\widetilde{\Delta}\right)$, and therefore, from Lemma 2, also

$$\max_{\tau'' \leq u \leq u_{\widetilde{\Delta}}(\tau'')} \left\|\theta_u - \tilde{\theta}\right\| \leq 2\widetilde{\Delta} < \rho. \tag{39}$$

Due to (35), the last inequality shows that

$$\theta_u \in B\left(\overline{\theta}, 2\rho\right), \ \tau'' \leq u \leq u_{\widetilde{\Delta}}(\tau''). \tag{40}$$

With (37), this yields the key result: $\theta_{u_{\tilde{\Delta}}(\tau'')} \in R_{\eta\Delta}$. Since $\theta_{\tau''}$ is a last value in $R_{\eta\Delta}$ before an entry in $S_{\frac{1}{2}\eta\Delta,}$ at time $\tau''+n''$, we must have $\tau''+n'' < u_{\tilde{\Delta}}(\tau'')$ whenever $\tau'' \geq \tau''\left(\tilde{\Delta}\right)$. For these τ'', it follows from (38), and (39) that $\|\theta_{\tau''} - \theta_{\tau''+n''}\| \leq 3\tilde{\Delta}$. Therefore, from (36), we have

$$\begin{aligned} V\left(\theta_{\tau''+n''}\right) &\leq V\left(\theta_{\tau''}\right) + \left|V\left(\theta_{\tau''}\right) - V\left(\theta_{\tau''+n''}\right)\right| \\ &\leq V\left(\overline{\theta}\right) - \eta\Delta + L_1 \left\|\theta_{\tau''} - \theta_{\tau''+n''}\right\| \\ &\leq V\left(\overline{\theta}\right) - \eta\Delta + 3\tilde{\Delta}L_1 \\ &< V\left(\overline{\theta}\right) - \frac{\eta}{2}\Delta, \end{aligned}$$

by virtue of (36). But this contradicts $\theta_{\tau''+n''} \in S_{\frac{1}{2}\eta\Delta}$. Thus the proofs of Lemma 5 and the Proposition are complete.

Remark 1. The gap in the proof given in DF is the lack of verification of (40), which seems to require a condition that forces $\left\|\theta - \overline{\theta}\right\|$ to be small when $\left|V(\theta) - V\left(\overline{\theta}\right)\right|$ is, as (31) does. No such condition is imposed in DF. When $d > 1$ and $\overline{\theta} \in \Theta_V$ is such that $\nabla V\left(\overline{\theta}\right) \neq 0$, the level sets $\left\{\theta \in \Theta_V : V(\theta) = V\left(\overline{\theta}\right)\right\}$ will be nonempty, so (31) will fail for every $m > 0$.

Remark 2. In the case $d = 1$, any antiderivative of $f(\theta)$ has the properties required of $V(\theta)$ in (b) of the Proposition, so (b) is not restrictive.

References

[1] Cantor, J. L. (2001) *Recursive and Batch Estimation for Misspecified ARMA Models.* Ph.D. Dissertation, Columbian School of Arts and Science, George Washington University.

[2] Derevitzkiĭ, D. P. and A. L. Fradkov (1981). *Applied Theory of Discrete Adaptive Control Systems.* (In Russian) Nauka: Moscow

[3] Solo, V. and X. Kong (1995). *Adaptive Signal Processing Algorithms: Stability and Performance.* Prentice Hall: Englewood Cliffs.

[4] Findley, D. F., B. M. Pötscher and C.-Z. Wei (2001). Modeling of Time Series Arrays by Multistep Prediction of Likelihood Methods. (submitted for publication).